

# Decomposition of Symmetry Using Cumulative Sub-Asymmetry Model for Square Contingency Tables

Kouji Tahata<sup>1</sup>, Tomoki Oyama and Sadao Tomizawa<sup>2</sup>

Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science, Noda City, Chiba, 278-8510, Japan

<sup>1</sup>kouji\_tahata@is.noda.tus.ac.jp; <sup>2</sup>tomizawa@is.noda.tus.ac.jp

## Abstract

For square contingency tables with ordered categories, the present paper gives the theorem that the symmetry model holds if and only if four models of (1) cumulative sub-asymmetry, (2) palindromic symmetry, (3) global symmetry, and (4) marginal means equality, hold simultaneously. It analyzes some data using this decomposition.

## Keywords

Ordinal Data; Palindromic Symmetry; Separation

## Introduction

Consider an  $r \times r$  square contingency table with ordered categories. Let  $p_{ij}$  denote the probability that an observation will fall in the  $i$ th row and  $j$ th column of the table ( $i = 1, \dots, r; j = 1, \dots, r$ ). Consider the symmetry (S) model, defined by

$$p_{ij} = p_{ji} \quad (i \neq j),$$

(Bowker, 1948; Bishop, Fienberg and Holland, 1975, p.282).

Let

$$G_{ij} = \sum_{s=1}^i \sum_{t=j}^r p_{st} \quad (i < j),$$

and

$$G_{ji} = \sum_{s=1}^i \sum_{t=j}^r p_{ts} \quad (i < j).$$

The S model can be expressed as

$$G_{ij} = G_{ji} \quad (i < j).$$

McCullagh (1978) considered the palindromic symmetry (PS) model, defined by

$$G_{ij} = \Delta \frac{\alpha_i}{\alpha_{j-1}} G_{ji} \quad (1 \leq i < j \leq r),$$

where  $\alpha_1 = 1$  without loss of generality. A special case of the PS model obtained by setting  $\Delta = 1$  and  $\alpha_1 = \dots = \alpha_{r-1}$  is the S model.

Let  $X$  and  $Y$  denote the row and the column variables, respectively. Consider the marginal means equality (ME) model, defined by  $E(X) = E(Y)$ , where

$$E(X) = \sum_{i=1}^r \sum_{j=1}^r i p_{ij},$$

and

$$E(Y) = \sum_{i=1}^r \sum_{j=1}^r jp_{ij}.$$

Also, the ME model can be expressed as

$$\sum_{i=1}^{r-1} G_{i,i+1} = \sum_{i=1}^{r-1} G_{i+1,i},$$

(see, Tahata, Yamamoto and Tomizawa, 2012).

As a model which indicates the structure of  $\{G_{ij}\}$  with  $|j-i|=2$ , namely,  $\{p_{ij}\}$  with  $|j-i|\geq 2$ , Tomizawa, Miyamoto and Ouchi (2006) proposed the cumulative sub-symmetry (CSS) model, defined by

$$G_{i,i+2} = G_{i+2,i} \quad (i = 1, 2, \dots, r-2).$$

Tahata et al. (2012) gave the theorem that the S model holds if and only if all the PS, ME and CSS models hold.

Tomizawa, Miyamoto, Yamamoto and Sugiyama (2007) considered the cumulative two ratios-parameter symmetry (C2RPS) model, defined by

$$\frac{G_{ij}}{G_{ji}} = \Gamma \Theta^{j-i} \quad (i < j).$$

The global symmetry (GS) model is defined by

$$\delta_U = \delta_L,$$

where

$$\delta_U = \sum_{i < j} p_{ij}, \quad \delta_L = \sum_{j < i} p_{ij},$$

(Read, 1977). Tahata, Yamamoto and Tomizawa (2013) gave the theorem that the S model holds if and only if all the C2RPS, GS and ME models hold.

The marginal homogeneity (MH) model is defined by

$$G_{i,i+1} = G_{i+1,i} \quad (i = 1, 2, \dots, r-1),$$

(Stuart, 1955). The model indicates that the marginal distribution for the row variable is equal to that for the column variable. Note that the model indicates the structure of  $\{G_{ij}\}$  with  $|j-i|=1$ . The extended MH model is defined by

$$G_{i,i+1} = \delta G_{i+1,i} \quad (i = 1, 2, \dots, r-1),$$

(Tomizawa, 1993). This model indicates the asymmetric structure of  $\{G_{ij}\}$  with  $|j-i|=1$ . Therefore, we are interested in considering the extended CSS model as the analogous to Tomizawa (1993). In this paper, we propose a model which indicates the structure of asymmetry for  $\{G_{ij}\}$  with  $|j-i|=2$ . Also, we show that the C2RPS model can be separated into the proposed model and the PS model. Using this result, we give the decomposition of the S model.

## Decompositions

This section proposes a model based on  $\{G_{ij}\}$ , and gives a decomposition of the C2RPS model.

Consider a model defined by

$$\frac{G_{i,i+2}}{G_{i+2,i}} = \Omega \quad (i = 1, 2, \dots, r-2).$$

We shall refer the model as the cumulative sub-asymmetry (CSAS) model. We now obtain the following theorem.

**Theorem 1.** The C2RPS model holds if and only if both the PS and CSAS models hold.

*Proof.* If the C2RPS model holds, then the PS model holds. Assume that the C2RPS model holds, and then we shall show that the CSAS model holds. We see from the C2RPS model that

$$\frac{G_{i,i+2}}{G_{i+2,i}} = \Gamma \Theta^{(i+2)-i} = \Gamma \Theta^2 \quad (i = 1, 2, \dots, r-2).$$

Thus the CSAS model holds.

Assume that both the PS and CSAS models hold, and then we shall show that the C2RPS model holds. We see from the PS model that

$$\frac{G_{i,i+2}}{G_{i+2,i}} = \Delta \frac{\alpha_i}{\alpha_{i+1}} \quad (i = 1, 2, \dots, r-2).$$

Then, we can set  $\alpha_1 = 1$  without loss of generality. Since the CSAS model holds,

$$\frac{G_{13}}{G_{31}} = \Delta \frac{\alpha_1}{\alpha_2} = \Delta \frac{1}{\alpha_2} = \Omega.$$

So, we obtain

$$\alpha_2 = \frac{\Delta}{\Omega}.$$

Similarly, we obtain  $\alpha_3 = (\Delta/\Omega)^2$  because

$$\frac{G_{24}}{G_{42}} = \Delta \frac{\alpha_2}{\alpha_3} = \frac{\Delta^2}{\Omega \alpha_3} = \Omega.$$

Therefore, we can see

$$\alpha_{i+1} = \left( \frac{\Delta}{\Omega} \right)^i \quad (i = 1, 2, \dots, r-2).$$

Thus

$$\frac{G_{ij}}{G_{ji}} = \Delta \frac{\alpha_i}{\alpha_{j-1}} = \frac{\Delta^2}{\Omega} \left( \frac{\Omega}{\Delta} \right)^{j-i} \quad (i < j).$$

The C2RPS model holds. The proof is completed.

We now obtain the following theorem from the result given by Tahata et al. (2013) and Theorem 1.

**Theorem 2.** The S model holds if and only if all the PS, CSAS, GS and ME models hold.

### Goodness-of-Fit Test

Let  $n_{ij}$  denote the observed frequency in the  $i$  th row and  $j$  th column of the  $r \times r$  table with  $n = \sum \sum n_{ij}$ , and let  $m_{ij}$  denote the corresponding expected frequency. Assume that a multinomial distribution applies to the  $r \times r$  table. We can test each model for goodness-of-fit by using the likelihood ratio chi-squared statistic  $G^2$  defined by

$$G^2 = 2 \sum_{i=1}^r \sum_{j=1}^r n_{ij} \log \left( \frac{n_{ij}}{\hat{m}_{ij}} \right),$$

where  $\hat{m}_{ij}$  is the maximum likelihood estimate of  $m_{ij}$  under the model. Table 1 presents the numbers of degrees of freedom for testing goodness-of-fit of the models.

TABLE 1 THE DEGREES OF FREEDOM OF MODELS FOR  $r \times r$  TABLE

| Models | Degrees of freedom |
|--------|--------------------|
| S      | $r(r-1)/2$         |
| C2RPS  | $r(r-1)/2 - 2$     |
| PS     | $(r-1)(r-2)/2$     |
| CSAS   | $r - 3$            |
| ME     | 1                  |
| GS     | 1                  |

### An Example

Table 2, taken directly from Tominaga (1979, p.130), is the cross-classification of Japanese father's and his son's academic background which were examined in 1955. Note that category (1) is elementary school; (2) junior high school; (3) high school; and (4) university. Table 3 gives the values of the likelihood ratio chi-square statistic  $G^2$  for models applied to these data. The CSAS model fits the data in Table 2 well.

TABLE 2 CROSS-CLASSIFICATION OF JAPANESE FATHER'S AND HIS SON'S ACADEMIC BACKGROUND; TAKEN FROM TOMINAGA (1979, P.130)

| Father's educational level | Son's educational level |                   |                   |                 | Total |
|----------------------------|-------------------------|-------------------|-------------------|-----------------|-------|
|                            | (1)                     | (2)               | (3)               | (4)             |       |
| (1)                        | 374<br>*(374.0000)      | 602<br>(602.0000) | 170<br>(167.3926) | 64<br>(64.6918) | 1210  |
| (2)                        | 18<br>(18.0000)         | 255<br>(255.0000) | 139<br>(139.0000) | 71<br>(72.9156) | 483   |
| (3)                        | 4<br>(6.9571)           | 23<br>(23.0000)   | 42<br>(42.0000)   | 55<br>(55.0000) | 124   |
| (4)                        | 2<br>(1.5482)           | 6<br>(3.4947)     | 17<br>(17.0000)   | 53<br>(53.0000) | 78    |
| Total                      | 398                     | 886               | 368               | 243             | 1895  |

Note: \*Estimated expected frequencies for the CSAS model.

Under the CSAS model, the maximum likelihood estimate of  $\Omega$  is  $\hat{\Omega} = 27.2872$ . Hence, under this model, the probability that a father's academic background status is  $i$  or below and his son's academic background status is  $i+2$  or above is estimated to be  $\hat{\Omega} = 27.2872$  times higher than the probability that the father's academic background status is  $i+2$  or above and his son's academic background status is  $i$  or below ( $i = 1, 2$ ).

Moreover, from Table 3, we see that the S model fits these data poorly, yielding  $G^2 = 1151.23$  with 6 degrees of freedom. Also, each of the PS, ME and GS models fit these data poorly, but the CSAS model fits these data well. By using the decomposition of the S model (i.e., Theorem 2), we shall consider the reason why the S model fits these data poorly. We can see that the poor fit of the S model is caused by the poor fit of the PS, ME and GS models rather than the CSAS model.

TABLE 3 LIKELIHOOD RATIO CHI-SQUARE VALUES  $G^2$  FOR MODELS APPLIED TO THE DATA IN TABLE 2

| Models | Degrees of freedom | $G^2$    |
|--------|--------------------|----------|
| S      | 6                  | 1151.23* |
| C2RPS  | 4                  | 31.20*   |
| PS     | 3                  | 31.13*   |
| CSAS   | 1                  | 3.18     |
| ME     | 1                  | 700.17*  |
| GS     | 1                  | 1093.22* |

Note: \*Significant at the 0.05 level.

## Concluding Remarks

We have proposed the CSAS model which indicates the structure of  $\{G_{ij}\}$  with  $|j-i|=2$ , namely,  $\{p_{ij}\}$  with  $|j-i|\geq 2$ . Using this model, we have given Theorem 1. The theorem may be useful to see the reason for the poor fit when the C2RPS model fits the data poorly. In practice, we can see from Theorem 1 that the poor fit of the C2RPS model is caused by the poor fit of the PS model rather than the CSAS model in example (see Table 3). For square contingency tables with ordered categories, Theorem 2 would be useful for seeing the reason for the poor fit when the S model fits the data poorly.

## ACKNOWLEDGMENT

The authors would like to thank an anonymous referee for helpful comments and suggestions.

## REFERENCES

- [1] Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975). *Discrete multivariate analysis: theory and practice*. MIT Press, Cambridge.
- [2] Bowker, A. H. (1948). A test for symmetry in contingency tables. *Journal of the American Statistical Association* **43**, 572-574.
- [3] McCullagh, P. (1978). A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika* **65**, 413-418.
- [4] Read, C. B. (1977). Partitioning chi-square in contingency tables: A teaching approach. *Communications in Statistics-Theory and Methods* **6**, 553-562.
- [5] Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika* **42**, 412-416.
- [6] Tahata, K., Yamamoto, K. and Tomizawa, S. (2012). Decomposition of symmetry using palindromic symmetry model in a two-way classification. *Journal of Statistics Applications and Probability* **1**, 201-204.
- [7] Tahata, K., Yamamoto, K. and Tomizawa, S. (2013). Decomposition of symmetry model into three models for cumulative probabilities in square contingency tables. *European Journal of Pure and Applied Mathematics* **6**, 299-306.
- [8] Tominaga, K. (1979). *Nippon no Kaisou Kouzou (Japanese Hierarchical Structure)*. University of Tokyo Press, Tokyo (in Japanese).
- [9] Tomizawa, S. (1993). A decomposition of the marginal homogeneity model for square contingency tables with ordered categories. *Calcutta Statistical Association Bulletin* **43**, 123-125.
- [10] Tomizawa, S., Miyamoto, N. and Ouchi, M. (2006). Decompositions of symmetry model into marginal homogeneity and distance subsymmetry in square contingency tables with ordered categories. *Revstat: Statistical Journal* **4**, 153-161.
- [11] Tomizawa, S., Miyamoto, N., Yamamoto, K. and Sugiyama, A. (2007). Extensions of linear diagonal-parameter symmetry and quasi-symmetry models for cumulative probabilities in square contingency tables. *Statistica Neerlandica* **61**, 273-283.